Citation for published version:
Johnson, SGB, Valenti, JJ & Keil, FC 2019, 'Simplicity and complexity preferences in causal explanation: An opponent heuristic account', *Cognitive Psychology*, vol. 113, 101222. https://doi.org/10.2139/ssrn.3194761, https://doi.org/10.1016/j.cogpsych.2019.05.004

DOI:
10.2139/ssrn.3194761
10.1016/j.cogpsych.2019.05.004

Publication date:
2019

Document Version
Peer reviewed version

Link to publication

Publisher Rights
CC BY-NC-ND

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

Simplicity and Complexity Preferences in Causal Explanation: An Opponent Heuristic Account

Samuel G. B. Johnson [1]
J. J. Valenti [2]
Frank C. Keil [3]

[1] University of Bath School of Management
[2] Johns Hopkins University, Department of Psychological and Brain Sciences
[3] Yale University, Department of Psychology

Corresponding Author: Sam Johnson

Address for Correspondence: The Avenue
Claverton Down
Bath BA2 7AY, Somerset
United Kingdom

Word Count: 9,709 (main text)

Last Revised: 9 April 2018 (revision submitted)

Electronic copy available at: https://ssrn.com/abstract=3194761

Abstract

People often prefer simple to complex explanations because they generally have higher prior probability. However, simpler explanations are not always normatively superior because they often do not account for the data as well as complex explanations. How do people negotiate this trade-off between prior probability (favoring simple explanations) and goodness-of-fit (favoring complex explanations)? Here, we argue that people use *opponent heuristics* to simplify this problem—that people use simplicity as a cue to prior probability but complexity as a cue to goodness-of-fit. Study 1 finds direct evidence for this claim. In subsequent studies, we examine factors that lead one or the other heuristic to predominate in a given context. Studies 2 and 3 find that people have a stronger simplicity preference in deterministic rather than stochastic contexts, while Studies 4 and 5 find that people have a stronger simplicity preference for physical rather than social causal systems, suggesting that people use abstract expectations about causal texture to modulate their explanatory inferences. Together, we argue that these cues and contextual moderators act as powerful constraints that can help to specify the otherwise ill-defined problem of what distributions to use in Bayesian hypothesis comparison.

Keywords: Causal reasoning; explanation; inference; understanding; simplicity.

Electronic copy available at: https://ssrn.com/abstract=3194761

The principle of parsimony has a long and venerable pedigree. It has been discussed since at least Aristotle, who wrote in his *Physics* that "nature operates in the shortest way possible," and it has since become one of the core tools in our argumentative arsenal as scientists (Sober, 1983). Of course, this principle was given its most famous formulation by William of Occam, who advised against "multiplying entities beyond necessity."

Simplicity is not only a core notion in science and philosophy, but may well be an organizing principle of cognition (Chater & Vitányi, 2003). People prefer simpler causal explanations (Lombrozo, 2007), category assignments (Pothos & Chater, 2002), and perceptual organizations (van der Helm & Leeuwenberg, 1996). Likewise, simpler concepts are more easily learned (Feldman, 2000), simpler items are easier to separate from noise (Hochberg & McAlister, 1953), and simplicity guides judgments of similarity (Hahn, Chater, & Richardson, 2003).

Simplicity has been defined in many different ways, including the number of assumptions, number of parameters free to vary in a model (Akaike, 1974), minimum description length (Rissanen, 1978; see also Kolmogorov, 1963), and the number of unexplained causes (Pacer & Lombrozo, 2017; Thagard, 1989). Here, we follow Lombrozo (2007) by operationalizing simplicity as the number of causes, with the assumption that our account is likely to generalize to other characterizations as well.

However defined, the principle of parsimony is not arbitrary. Other things equal, simpler theories or explanations are more likely to be true because they have higher prior probability under many conditions. For example, imagine you hear about an airplane crash. Suppose that there are two possible explanations—either a failure of the landing system (A), or a failure of both the wings (B) and the engine (C). Most people would consider explanation A more satisfying, because it involves only one cause (Lombrozo, 2007). This reasoning is normative if the causes are independent and have similar prior probabilities: If the cause of each mechanical failure is 1 in 1000, then explanation A has a probability of 1/1000 *a priori*, whereas explanation $\{B,C\}$ has a probability of only 1/1,000,000.

This intuition is captured by Bayes' theorem, which can be used to compare the relative probability of two explanations given some data. The posterior odds favoring explanation A over explanation $\{B,C\}$ are equal to:

$$\frac{P(A|Crash)}{P(B,C|Crash)} = \frac{P(A)}{P(B,C)} \cdot \frac{P(Crash|A)}{P(Crash|B,C)} = \frac{.001}{.001 \cdot .001} \cdot \frac{1}{1} = \frac{1000}{1}$$

That is, after observing the data (*Crash*), the odds favoring A over $\{B,C\}$ are equal to the prior odds favoring A over $\{B,C\}$ before observing any data [$P(A)$ / $P(B,C)$], multiplied by the likelihood ratio, or fit of each explanation to the data [$P(Crash|A)$ / $P(Crash|B,C)$]. Assuming that either explanation would lead deterministically to a plane crash (so that $P(Crash|A) = P(Crash|B,C) = 1$, and $P(Crash|A)$ / $P(Crash|B,C) = 1$), the posterior odds are determined only by the prior odds, and favor the simpler explanation A by a factor of 1000.

Consistent with this analysis, Lombrozo (2007) found that people use simplicity as a heuristic for estimating prior probabilities. In her experiments, participants performing simulated medical

Electronic copy available at: https://ssrn.com/abstract=3194761

diagnoses would not accept a complex explanation over a simple one unless the prior probabilities favored the complex explanation by at least a factor of 4. Further, participants who had a simplicity bias had distorted memories of the disease base rates, recalling the simpler explanations as having had higher prior probabilities than they in fact did. Thus, people's preference for simple explanations, though sometimes stronger than normatively warranted, appears to track the probabilistic logic favoring simpler explanations.

Yet, intuition does not always seem so clear about the virtue of simplicity. As Oscar Wilde's Algernon noted, "The truth is rarely pure and never simple" (Wilde, 2004/1895). That is to say, simplicity has its limits: The facts are not always simple enough to warrant a simple explanation. In statistical terms, a simple and a complex explanation do not always fit the data equally well. For instance, contrast again explanation A (landing system failure) and $\{B,C\}$ (wing and engine failure) for the airplane crash. Imagine we also observe black smoke coming out of the airplane's engine prior to the crash. In that case, the likelihoods for A and $\{B,C\}$ are not equal because the more complex explanation $\{B,C\}$ can account for more of the data (both crash and the black smoke) than explanation A (which explains the crash but not the black smoke). Imagine that the totality of the data (crash plus smoke) would occur with only 1/10,000 probability given explanation A, but would occur with probability 1 given explanation $\{B,C\}$. That is, the posterior odds are now:

$$\frac{P(A|Crash)}{P(B,C|Crash)} = \frac{P(A)}{P(B,C)} \cdot \frac{P(Crash|A)}{P(Crash|B,C)} = \frac{.001}{.001 \cdot .001} \cdot \frac{.0001}{1} = \frac{1}{10}$$

These odds actually favor the more *complex* explanation. Indeed, complex explanations generally allow more opportunities to explain the data because they invoke more causes or degrees of freedom (Forster & Sober, 1994).

Opponent Heuristics

Given these considerations, it is implausible that people invariably favor arbitrarily simple hypotheses. Despite the enthusiasm of Lombrozo's (2007) participants for simple explanations, there must be boundary conditions on this simplicity bias, for two reasons. First, simplicity concerns only the form of the *hypothesis*, not the nature of the observations. Since the central goal of explanatory reasoning is to account for the data, simplicity must be coordinated with other cues and mechanisms for assessing an explanation. This will inevitably involve some factors that can potentially lead a reasoner to adopt a more complex explanation. Second, there is generally a U-shaped curve in how simple an explanation ought to be. Too complex, and the explanation has a lower prior probability and overfits the data; too simple, and it does not account for the nuance of the phenomenon (Forster & Sober, 1994). How, if at all, does cognition perform this trade-off?

This article presents evidence for an *opponent heuristic* theory of simplicity and complexity preferences. This view incorporates Lombrozo's (2007) insight that people use simplicity to estimate prior probability—the $P(H_i)$ terms in Bayesian hypothesis comparison—but couples it with the idea that people also use *complexity* to estimate likelihoods—the $P(E|H_i)$ terms that measure the goodness-of-fit of the evidence to the data.

Electronic copy available at: https://ssrn.com/abstract=3194761

For example, if a patient is sneezing and has a stomach ache, one explanation could be that the patient has a cold. This explanation is simple, but is an imperfect fit to the data. That is, if we took a random sample of the population, a reasonably large fraction of these people would have a cold at any given time—so this explanation has high prior probability. But among those people who *have* a cold, how many of them would both be sneezing and have a stomach ache? The *facts* here are complex, and this simple explanation does not fit very well.

In contrast, the patient could have both allergies and a stomach virus. This explanation is more complex, but fits the data neatly. That is, in a random sample of the population, a fairly small number would have both allergies and a stomach virus. Yet, many of those who *do* have both diseases would likely be suffering from both sneezing and a stomach ache. Even though the prior probability of this complex explanation is low, it fits the data very well.

In this case, simplicity seems to be associated with our estimate of prior probability and complexity seems to be associated with our estimate of likelihood. Of course, this explanation was engineered to produce these intuitions by relying on specific beliefs we have about these diseases. The opponent heuristic account proposes that people also use simplicity and complexity as cues in cases where they cannot estimate probabilities directly from background knowledge.

Initial evidence for this opponent heuristic proposal comes from studies of intuitive curve-fitting—a superficially dissimilar but deeply related problem to causal explanation. For any set of scatterplot data, many different trend curves can be drawn to explain the data, but modern statistical theory can tell us exactly which curve has the best predictive power, fitting as much of the underlying signal as possible while fitting little of the noise (e.g., Akaike, 1974). Laypeople, however, tend to choose curves that are more *complex* than they normatively should be, rather than curves that are too simple (Johnson, Jin, & Keil, 2014), as one would expect if people only have a simplicity heuristic but no complexity heuristic. Indeed, these curve-fitting studies uncovered direct evidence of a complexity bias, because participants believed that more complex were literally closer fits to the data, even when the actual fit was the same.

Why is this pair of heuristics useful? It would seem that simplicity is just the *absence* of complexity. How, then, can a *pair* of heuristics accomplish any more than a single heuristic, when these two heuristics rely on the same cue? To put the point differently, why do we not instead assume that people seek the level of simplicity necessary to maximize $P(H)P(E|H)$, perhaps by using a single heuristic cue in a U-shaped manner, increasing complexity up to a point but no further?

To see the advantage of opponent heuristics, consider that organisms face a trade-off between the flexibility of a problem-solving technique and its computational complexity. For several reasons, humans face limits in the kinds of computations we can execute. In the explanatory domain, we face informational limits because we do not have access to all relevant evidence (Johnson, Rajeev-Kumar, & Keil, 2016); we face specification limits because we often lack a principled way of assigning point probabilities to possibilities (Knight, 1921); and we face cognitive limits because we have (sharply) finite memory and must avoid exponential explosions. Yet, we also face a constantly changing environment and must adapt to new information and novel situations. How might such an organism best cope with these challenges?

Electronic copy available at: https://ssrn.com/abstract=3194761

What we *cannot* do is to optimize—to write $\mathrm{P}(H)\mathrm{P}(E|H)$ as a function of complexity k, set the first derivative to zero, and solve for k. Such a direct method faces the gamut of problems just noted—we cannot hope to enumerate all the relevant factors (information limits), cannot write this expression in a precise way (specification limits), and even if we could overcome these problems, the resulting problem would be too cumbersome for a human cognitive agent to solve. Indeed, these are the same reasons that artificial intelligence systems have not yet made significant progress toward such kinds of high-level cognitive tasks. The optimization method is highly flexible, but it has too many moving parts and becomes impossible to execute.

Yet, we cannot simply take a prefabricated solution off the shelf either. An explanation invoking one cause will often be too simple (it does not fit the data well), but an explanation invoking ten causes will often be too complex (it has a very low prior). Moreover, we cannot simply assume that the optimal solution to this trade-off is some particular intermediate degree of complexity—for instance, to assume that all happenings in the universe have three causes, no more and no less. Such a method is hopelessly rigid. It would be maladaptive in the literal sense of the word, since an organism following such a rule would be unable to adjust its beliefs to changes in the circumstances, such as the nature of the causal system under investigation (e.g., the degree of determinism and content domain, as we explore in our studies). This prefabrication approach is easy to execute, but has too *few* moving parts and is overly restrictive.

A pair of opponent heuristics has the potential to outperform both approaches. Suppose that the reasoner begins by considering priors, using the simplicity heuristic. Depending on the context (see below), one might have a stronger or weaker preference for the simpler explanation. For example, if a reasoner believes that a particular explanatory domain typically has more complex causal connections, she may assign lower weight to simplicity as a cue for prior probability when evaluating explanations in that domain. Next, the reasoner would adjust this estimate to take account of the evidence, using the complexity heuristic. Once again, depending on the context, this might be a larger or smaller adjustment toward the complex explanation. (In practice, these steps might be executed simultaneously, akin to averaging, rather than sequentially.)

A procedure of this kind allows the reasoner to take account of the way that the context of the problem might independently influence the prior and the likelihood (outperforming the prefabrication method in flexibility), but does not require the computational complexity of the optimization method. Opponent heuristics restrain reasoners so that they do not favor the simplest explanations in all cases, as they would if people used a simplicity heuristic alone as a cue to prior probability. At the same time, they allow simplicity to take a high priority when the context demands it (e.g., for physical rather than for social systems) by using smart yet fallible rules-of-thumb rather than intractable calculations. Although there is no reason to think that a context-sensitive dual heuristic solution will give an optimal answer, there *is* reason to think that it may bring the reasoner closer to the right part of the hypothesis space, compared to the inflexible prefabrication method or to either heuristic working alone.

Electronic copy available at: https://ssrn.com/abstract=3194761

Contextual Factors

Although there may be several kinds of factors that modulate explanatory inferences and influence the relative weight of the simplicity and complexity heuristics, two particularly likely candidates—both considered empirically here—are the *determinism* and the *content domain* of the causal system.

Determinism. In previous studies of simplicity (Lombrozo, 2007; see also Bonawitz & Lombrozo, 2012 in children), explanations have been produced for deterministic causal systems. In such systems, it is *rational* to prefer simple explanations, so long as each piece of evidence has one causal factor in the explanation. For example, consider the causal structure depicted in Figure 1. If disease *A* *always* causes symptoms *X* and *Y*, while disease *B* *always* causes symptom *X* and disease *C* *always* causes symptom *Y*, the issue of likelihoods or goodness-of-fit simply does not come up: Disease *A* *perfectly* explains the evidence, and so do Diseases *B* and *C* together. The only issue is which explanation has the higher prior probability, and the simplicity heuristic tells us that, absent any other information, the answer is Disease *A*. Therefore, there is no reason to invoke a complexity heuristic to countervail against the presumption of a simple explanation, leading to a strong simplicity bias.

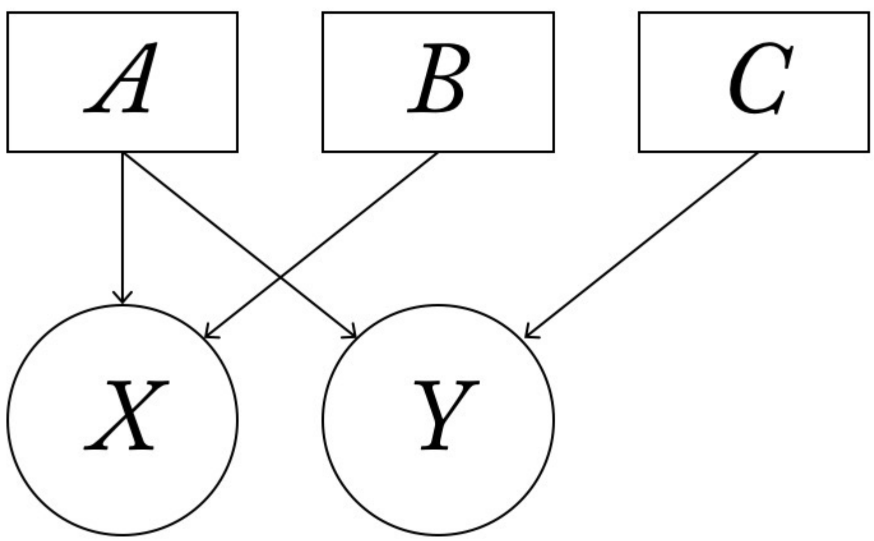

Figure 1. Causal structure where simple and complex explanations are in competition.

Note. The *X* and *Y* nodes designate the observed evidence, which can be explained either by cause *A* alone, or by the combination of causes *B* and *C*.

In contrast, when the causal system is stochastic, the likelihoods become a more crucial part of the computation. If disease *A* sometimes causes *X* and sometimes causes *Y*, while disease *B* sometimes causes *X* and disease *C* sometimes causes *Y*, it is difficult to evaluate whether the evidence (symptoms *X* and *Y*) are made likelier by disease *A* or by diseases *B* and *C* combined: It

Electronic copy available at: https://ssrn.com/abstract=3194761

depends critically on the nature of "sometimes." Yet, in the real world, it is the exception rather than the rule to have precise quantitative information about these likelihoods in stochastic systems. If people rely on a complexity heuristic in such cases, they would judge the likelihood of the evidence to be higher for an individual with two diseases than for an individual with one disease. Thus, the complexity heuristic would be likely to be invoked in a stochastic context, leading to a weaker bias for the simple explanation (or perhaps even a bias for the complex explanation, as found in Johnson et al., 2014).

Domain. Even for a novel problem, people rarely approach inference as a blank slate. We have generalized knowledge about various content domains, including social, biological, and physical causal systems (Wellman & Gelman, 1992). This knowledge can take the form of intuitive theories about specific causal patterns such as human choice (Johnson & Rips, 2015), inheritance (Springer & Keil, 1989), or gravity (Hood, 1995), but it can also take the form of more abstract expectations.

Most important for current purposes, people seem to have different beliefs about the causal textures of different content domains. Whereas people tend to identify physical events as having relatively few causes, social events are often thought to have many causes (Strickland, Silver, & Keil, 2017). This suggests that people may calibrate their prior expectations to more complex explanations in the social domain, compared to the physical domain. Furthermore, people may even deploy different causal concepts across domains (Lombrozo, 2010). Whereas causal claims about physical systems appear to be evaluated in terms of transference and contact (e.g., Dowe, 2000), social causal claims appear to be evaluated counterfactually (e.g., Lewis, 2000; Mackie, 1965). This too may reinforce the intuition that physical events typically result from highly specified causal factors, whereas social events result from more complex configurations of counterfactual conditions. Since such complex conditions can seldom be known fully, social systems are often highly unpredictable.

As a consequence of these domain-specific expectations, people may rely on simplicity as a cue to prior probability to a differing degree across domains. Whereas simplicity is likely to be a potent heuristic for evaluating explanations of physical causation, it may be a weaker cue for evaluating explanations of social causation, if people have a meta-theory that assigns higher prior probabilities to complex social causal explanations, as compared to physical causal explanations. In addition, if social causal systems are seen as more stochastic, this would increase the importance of the complexity heuristic for evaluating explanations of social causation, as compared to physical causation. With a weaker simplicity heuristic and stronger complexity heuristic, people may therefore have a smaller bias toward simple explanations in the social domain.

Empirical Approach

The studies reported here test the opponent heuristic account and examine contextual factors that can influence the adoption of simple and complex explanations. First, Study 1 tests these heuristics directly, using artificial stimuli modeled closely on Lombrozo's (2007) items. This study attempts to (a) test more directly Lombrozo's notion that people assign higher prior probabilities to simple explanations and (b) test the novel claim here that people assign higher likelihoods to

Electronic copy available at: https://ssrn.com/abstract=3194761

complex explanations. This would establish the fundamental mechanics of the opponent heuristic account.

Second, Studies 2 and 3 test the proposal that these heuristics are weighted differentially in stochastic versus deterministic causal systems. This follows from the idea that complexity is used for assessing likelihoods, but in deterministic systems with perfect likelihoods, such a heuristic is unnecessary, leading to a stronger simplicity bias. Study 2 tests this possibility by asking participants to compare simple and complex explanations when the stochasticity of the system varies. Study 3 builds on these findings by asking participants to rate 1-cause, 2-cause, and 3-cause explanations on separate scales.

Third, Studies 4 and 5 test the possibility that people favor different levels of complexity across domains. Study 4 asks participants directly to compare priors and likelihoods for simple and complex explanations, across physical, biological, social, and artifact systems. The opponent heuristic account predicts a stronger complexity preference in assessing likelihoods, compared to priors, but it is consistent with the further prediction that the magnitude of explanatory preferences differ across domains (i.e., a stronger simplicity preference for physical systems than for social systems). Study 5 asks participants once again to rate explanations directly, varying both stochasticity and domain. This allows the evaluation of both contextual moderators in the same design.

In addition to the studies reported in the main text, we have conducted several other studies (reported in the online Supplementary Materials in the interest of brevity). These studies replicate the primary findings reported in the main text under a variety of different experimental conditions (e.g., unipolar versus bipolar scales; different specifications of probabilities), and examine a number of ancillary questions raised by the primary studies.

Study 1

To a Bayesian, the key quantities required to compare two hypotheses are the relative prior probabilities of the hypotheses (the *prior odds*), and the relative fit of each explanation to the data (the *likelihood ratio*). Study 1 tests whether people use simplicity to estimate these quantities.

Study 1A seeks converging evidence for Lombrozo's (2007) claim that people assign higher prior probabilities to simple hypotheses. Study 1B tests whether this heuristic favoring simple explanations might be opposed by a heuristic that assigns higher likelihoods to more *complex* explanations: Do people believe that complex explanations are better fits to the data?

Lombrozo's (2007) study used artificial stimuli to isolate participants' background knowledge, asking participants to make explanatory inferences for a single item, diagnosing the potential diseases of an alien. Here, we retain Lombrozo's artificial stimuli to facilitate comparison, but generalize these effects well beyond aliens—to elves, centaurs, and mermaids. (The skeptical reader will have to wait until Study 4 for more realistic stimuli across domains.)

Methods. Participants in all studies were recruited from Amazon Mechanical Turk. Each study included a series of memory check questions at the end, to assure attentiveness to the stimuli, and participants were excluded from analysis if they answered more than 33% incorrectly.

Electronic copy available at: https://ssrn.com/abstract=3194761

Participants (N = 80, 9 excluded) were randomly assigned to Study 1A (and asked to make judgments about prior probabilities) or to Study 1B (and asked to make judgments about likelihoods). In both studies, participants completed four items similar to the following problem, each item with a different cover story (different species of creature, names, symptoms, etc.):

There is a population of elves that lives at Gelfert's Glacier. Sometimes the elves have medical problems such as feverish muffets or wrinkled ears.

A **Yewlie infection** can cause feverish muffets.
A **Yewlie infection** can cause wrinkled ears.
Hepz's disease can cause feverish muffets.
Aeona's syndrome can cause wrinkled ears.

Nothing else is known to cause an elf's muffets to be feverish or the development of wrinkled ears.

On the same screen as this information, participants completed a series of 10 true/false questions (e.g., "Aeona's syndrome can cause wrinkled ears") to ensure comprehension.

Participants in Study 1A were then asked to judge the relative prior probabilities ("Imagine that we randomly select an elf from Gelfert's Glacier. Which of the following types of elves do you think we are more likely to have selected?") on a continuous scale from –5 to 5 with one end corresponding to the simple explanation ("An elf who has a Yewlie infection only") and one end corresponding to the complex explanation ("An elf who has both Hepz's disease and Aeona's syndrome"). The scale was oriented randomly across items. Participants in Study 1B were asked to judge the relative likelihoods ("Imagine an elf who has a Yewlie infection only, and another elf who has both Hepz's disease and Aeona's syndrome. Which elf do you think is more likely to develop both feverish muffets and wrinkled ears?") on the same scale. Items were completed in a random order.

Results and discussion. In data analysis for all studies, scores were coded so that negative numbers correspond to the simple explanation and positive numbers to the complex explanation.

As shown in Table 1, participants in Study 1A used a simplicity heuristic, indicating that a randomly selected elf was more likely to have one disease than two diseases [$t(33) = 7.19$, $p < .001$, $d = 1.23$]. This is consistent with Lombrozo's studies, where overwhelming prior odds (e.g., 4 to 1) were required before participants would favor a complex over a simple explanation in deterministic cases (Lombrozo, 2007, Experiment 2) and where participants misremembered the prior probabilities of simple causal explanations as higher than they actually were.

Quantity	Judgment
Prior Odds (Study 1A)	–2.19 (1.78)
Likelihood Ratio (Study 1B)	1.41 (2.35)

Table 1. Results of Study 1.

Electronic copy available at: https://ssrn.com/abstract=3194761

Note. Entries are relative probability judgments. Negative scores correspond to simple explanations, and positive scores to complex explanations. Scale ranges from –5 to 5. (SDs in parentheses.)

However, the story was different for judgments of likelihoods or goodness-of-fit. Here, participants favored the *complex* explanation [$t(36) = 3.65$, $p = .001$, $d = 0.60$], as shown in Table 1. This complexity bias in estimating likelihoods was substantial in magnitude ($d = 0.60$), though smaller than the simplicity bias in estimating priors ($d = 1.23$), at least for these stimuli.

Study S1 in the Supplementary Materials replicates the basic finding of Study 1 with two changes: (a) The vignette text specified that the priors and likelihoods are to be evaluated relative to a specific elf's symptoms rather than to elves in general, and (b) these judgments were made on separate scales for the simple and complex explanations. As in Study 1, the prior was judged higher for simple than complex explanations (the simplicity heuristic), the likelihood of the data was judged higher for the complex than the simple explanation (the complexity heuristic), and the simplicity heuristic was stronger than the complexity heuristic.

These results shows that people do not blindly prefer simple explanations, but instead calibrate their preferences according to the question asked. Even though the problem did not include any information about prior probabilities or likelihoods, participants used simplicity and complexity to estimate these quantities in opposing ways. Studies 2–5 look at ways that these heuristics inform explanatory judgments and how their use is shaped by contextual factors.

Study 2

The simplicity and complexity heuristics rely on the same *cue* (simplicity and complexity, both defined operationally as the number of causes) to estimate different *quantities* (the prior odds and likelihood ratio, respectively). These opponent heuristics cannot reach a solution that is both unique (for a given situation) and flexible (potentially differing across contexts) without additional assumptions about the way that the strengths of these heuristics are modulated across contexts. Studies 2 and 3 look at the probabilistic structure of the causal system as one contextual moderator, and Studies 4 and 5 look at content domain as a second moderator.

In any causal system where there is uncertainty about which explanation is correct, the prior probabilities of each explanation must be less than 1, since otherwise there is no reason to observe any data (as it will fail to move the posteriors). However, the *likelihoods* differ across deterministic and stochastic systems. In deterministic systems, the evidence is always produced with probability 1 by its causes, whereas in stochastic systems, these likelihoods are less than 1.

If explanatory heuristics exist in part because degrees of uncertainty are difficult to estimate and to use in computations, then a simplicity heuristic will always be a useful tool for estimating priors, since they are always uncertain. However, a complexity heuristic is only useful in stochastic systems, where the likelihoods are uncertain. Thus, both heuristics should be at work in stochastic systems (a simplicity heuristic pushing toward simpler explanations and a complexity heuristic pushing toward more complex explanations), whereas only the simplicity heuristic applies in deterministic systems (pushing toward simpler explanations, without an opposing force pushing toward more complex

Electronic copy available at: https://ssrn.com/abstract=3194761

explanations). This leads to the prediction that people should especially favor simple explanations for deterministic systems.

Methods. Participants (N = 80, 14 excluded) completed four items corresponding to the cover stories used in Study 1. For one of these items—in the 100% condition—the causal system was described as deterministic, in that the diseases always led to their symptoms (100% likelihood):

There is a population of 750 aliens that lives on planet Zorg. You are a doctor trying to understand an alien's medical problem. The alien, Treda, has two symptoms: Treda's minttels are sore and Treda has developed purple spots.

Tritchet's syndrome always (*100% of the time*) causes both sore minttels and purple spots.
Morad's disease always (*100% of the time*) causes sore minttels, but the disease never (*0% of the time*) causes purple spots.
When an alien has a **Humel infection**, that alien will always (*100% of the time*) develop purple spots, but the infection will never (*0% of the time*) cause sore minttels.

Nothing else is known to cause an alien's minttels to be sore or the development of purple spots.

The other three items corresponded to the 90%, 80%, and 70% conditions, which differed only in the causal system being described as stochastic:

Tritchet's syndrome often ([*80/65/50]% of the time*) causes both sore minttels and purple spots.
Morad's disease often (([*90/80/70]% of the time*) causes sore minttels, but the disease never (*0% of the time*) causes purple spots.
When an alien has a **Humel infection**, that alien will often (([*90/80/70]% of the time*) develop purple spots, but the infection will never (*0% of the time*) cause sore minttels.

After reading this information, participants were asked about their favored explanation ("Which do you think is the most satisfying explanation for Treda's symptoms?") on a scale from 0 (the simple explanation) to 10 (the complex explanation). The conditions were balanced across the cover stories using a Latin square, and items were completed in a random order.

Results and discussion. As shown in Table 2, participants strongly preferred the simple explanation [$t(65) = 15.84$, $p < .001$, $d = -1.95$] given deterministic (100%) likelihoods. This is a near exact replication of Lombrozo's (2007, Experiment 1) finding that people have an extremely strong preference for simple explanations in deterministic causal systems.

The new question addressed here is whether this preference would differ in the stochastic conditions, where a complexity heuristic would be more likely at play for understanding the likelihoods. To keep the likelihood ratio objectively identical across conditions, the likelihood for the simple explanation must equal the product of the likelihoods for the components of the complex explanation (i.e., 90% × 90% ≈ 80%, 80% × 80% ≈ 65%, and 70% × 70% ≈ 50%). This calculation assumes that people believe diseases to cause their symptoms independently. We did not ask

Electronic copy available at: https://ssrn.com/abstract=3194761

participants about this in the current study. However, we did ask about this in Study S2 in the Supplementary Material, excluding participants who denied this independence.

As predicted by the opponent heuristic account, the simplicity bias was weaker in each of the three stochastic conditions (see Table 2), although participants still had a robust simplicity preference in each of the three conditions [$t(65) = 9.09$, $p < .001$, $d = 1.12$ for the 90% condition; $t(65) = 7.86$, $p < .001$, $d = 0.97$ for the 80% condition, $t(65) = 8.24$, $p < .001$, $d = 1.01$ for the 70% condition]. The simplicity bias in the stochastic conditions, while large (with Cohen's d varying from 0.97 to 1.12), are considerably smaller than the bias in the deterministic condition ($d = 1.95$), consistent with predictions.

Likelihood	Judgment
100% (Deterministic)	–3.81 (1.95)
90% (Stochastic)	–3.00 (2.68)
80% (Stochastic)	–2.50 (2.58)
70% (Stochastic)	–2.48 (2.45)

Table 2. Results of Study 2.

Note. Entries are explanatory judgments. Negative scores correspond to simple explanations, and positive scores to complex explanations. Scale ranges from –5 to 5. (SDs in parentheses.)

However, this design is subject to concerns about demand characteristics and difficulties with probabilities that are unrelated to the proposed mechanisms. In particular, the deterministic condition set all likelihoods to 100%, whereas the stochastic condition had to set different likelihoods for the simple explanation and for each component of the complex explanation, where the numerical likelihood for the simple explanation was lower, in order to equalize the actual likelihoods. Could people have relied on a very simple strategy, such as comparing these numerical likelihoods (100% vs. 100% and 90% vs. 80% for complex vs. simple, respectively), favoring the complex explanation more in the stochastic conditions merely because it was superficially associated with higher numbers?

If this were the case, people should be increasingly less biased toward the simple explanation as the difference between the simple and complex likelihoods increased. This difference increases not only between the deterministic and stochastic conditions, but also *across* the stochastic conditions (90% vs. 80%, 80% vs. 65%, and 70% vs. 50%). Thus, on this deflationary account there should be large gaps not only between the deterministic and stochastic conditions, but also among the stochastic conditions. In contrast, the opponent heuristic account predicts a qualitative shift between the deterministic condition, where the likelihoods involve no uncertainty, and the stochastic conditions, which introduce uncertainty.

The data are more consistent with the latter prediction, as suggested by the similar effect sizes of the simplicity bias across the three stochastic conditions. Further, there is a significant difference between the 100% and 90% conditions, where we shift from deterministic to stochastic [$t(65) =$

Electronic copy available at: https://ssrn.com/abstract=3194761

2.61, p = .011, d = 0.32]. However, the difference between the 90% and 80% conditions reaches only marginal significance [t(65) = 1.88, p = .064, d = 0.23] and the difference between the 80% and 70% conditions is nowhere near significant [t(65) = 0.04, p = .97, d = 0.01]. The deflationary account would predict equally large differences across these sets of conditions.

Thus, determinism may play a role in striking the balance between the simplicity and complexity heuristics. These results also resolve a puzzle about Lombrozo's (2007) findings. Given that people are reasonably well-calibrated in evaluating explanations in the real world, it is surprising to see such a striking simplicity bias as one finds in her studies, with prior odds of 4-to-1 required to override a simplicity preference when the evidence is perfectly consistent with either hypothesis. Study 2 found that in more ecologically realistic conditions, where the evidence is not perfectly predicted by any explanation, people are more likely to hedge their bets, revealing use of a complexity heuristic. Thus, people may make more accurate explanatory inferences in realistic, stochastic environments, albeit still with a strong preference for simplicity.

Study 3

Whereas Studies 1 and 2 asked participants to *compare* two potential explanations, Study 3 asks participants to rate the probability of each explanation independently, with a greater range of complexity than in previous studies.

Methods. Participants (N = 159, 60 excluded) completed four items, using modified versions of the cover stories from Studies 1 and 2. Participants were randomly assigned to read about either deterministic or stochastic causal systems. In the deterministic condition, participants read four items similar to the following:

You are a doctor trying to understand an elf's medical problem. You are deciding on his diagnosis.

The elf, Wenlie, has three symptoms: Wenlie has feverish muffets, Wenlie's ears have wrinkled up, and Wenlie has a Nurino deficiency.

A **Yewlie Infection** *always* causes an elf to develop feverish muffets, wrinkled ears, and a Nurino deficiency.

Hepz's Disease *always* causes feverish muffets and wrinkled ears.
Aeona's Syndrome *always* causes an elf to have deficient Nurinos.

An elf with **Jonjo's Disease** *always* develops feverish muffets.
McArdel's Disease *always* causes wrinkled ears.
A **Jeong Infection** *always* causes a Nurino Deficiency.

Nothing else is known to cause an elf's muffets to be feverish, the development of wrinkled ears, or a Nurino deficiency.

The stochastic condition differed only in replacing the word "always" with the word "sometimes" in all causal statements. For half of participants, the explanations were described in

Electronic copy available at: https://ssrn.com/abstract=3194761

order from simplest to most complex (and the ratings made in that order) and for half of participants the converse order was used.

After reading this information, participants were asked to assign probabilities to causal explanations of varying complexity ("Please estimate the probability of Wenlie having each combination of diseases. Please ensure that all three options add up to 100%."). Participants then estimated the probabilities for "Wenlie has a Yewlie Infection," "Wenlie has Hepz's Disease and Aeona's Syndrome," and "Wenlie has Jonjo's Disease, McArdel's Disease, and a Jeong Infection"—the three explanations that account for all the observed symptoms. These ratings were made on a scale from 0% to 100% for each explanation, and participants were instructed to ensure the probabilities sum to 100%. Participants whose sum (averaged across the four problems) was not between 80% and 120% were excluded from analysis. (Studies S2 and S3 in the Supplementary Materials did not ask participants to sum the probabilities to 100%, and produced similar results.) Items were completed in a random order.

Results and discussion. Table 3 lists the mean posterior probabilities assigned to each explanation in each condition, as well as a benchmark posterior to facilitate statistical analysis. This benchmark posterior simply assumes that the probability of the 2-cause explanation is the squared probability of the 1-cause explanation and the probability of the 3-cause explanation is the cubed probability of the 1-cause explanation, with the probabilities summing to 1. This is not necessarily normative, since the explanations need not be mutually exclusive or exhaustive. However, they reflect a nonarbitrary reference point to which we can compare our participants' responses. Although this reference point seems appropriate since it flows from one set of plausible assumptions participants could make, the results are not dependent on the precise point chosen.

Relative to this benchmark, participants assigned greater weight to the simple (1-cause) explanation, both in the deterministic [$t(43) = 4.31$, $p < .001$, $d = 0.65$] and the stochastic conditions [$t(54) = 3.46$, $p = .001$, $d = 0.47$]. This led participants, correspondingly, to place less weight on the more complex 2-cause and 3-cause explanations, compared to their benchmarks. This greater weight on the simple explanation is consistent with the overall simplicity bias we have been finding in our previous studies (as well as Lombrozo, 2007).

Most critically, the distributions differed across conditions. To test this, we computed the "Euclidean deviance" relative to the benchmark for each participant, by taking the deviation between each participant's average judgment of each explanation and the normative benchmark, squaring these deviations, summing the squares, and computing the square root. (This is the same formula used to compute distances between two coordinates in a three-dimensional space.) This is a simple measure of how much a probability distribution differs from a particular fixed point, and differences between the deterministic and stochastic conditions would imply that those conditions also differ from one another. These scores were significantly larger in the deterministic condition than in the stochastic condition [$t(97) = 2.04$, $p = .044$, $d = 0.41$], indicating that people diverged significantly more from the benchmark for stochastic systems, shifting toward more complex explanations in the stochastic condition.

Electronic copy available at: https://ssrn.com/abstract=3194761

	Posterior Probabilities		
Explanation	**Deterministic**	**Stochastic**	**Benchmark**
1-cause	67.9% (20.8%)	62.6% (17.7%)	54.4%
2-cause	20.8% (11.9%)	24.4% (11.3%)	29.6%
3-cause	12.6% (10.8%)	13.6% (9.1%)	16.1%
Euclidean Deviance	27.0% (14.8%)	21.4% (12.7%)	

Table 3. Results of Study 3.

Note. Entries for the Deterministic and Stochastic columns are judged posterior probabilities, expressed as percentages (SDs in parentheses). Entries in the Benchmark column were calculated as described in the main text. The Euclidean Deviance row gives the average Euclidean distance among participants in each condition to the benchmark response vector.

In fact, these results may well underestimate the true difference in inferences between stochastic and deterministic environments. This is because the likelihoods for all three explanations are normatively equal (to 100%) in the deterministic condition, but the objective likelihoods may actually favor the simple explanation in the stochastic condition, if participants assume that a given probability word (e.g., "sometimes") always refers to the same probability. For instance, suppose that participants assume that "sometimes" means 90%. Then the likelihood of the evidence is 0.90 given the 1-cause explanation, 0.81 given the 2-cause explanation (0.9×0.9), and 0.729 given the 3-cause explanation ($0.9 \times 0.9 \times 0.9$). Thus, the likelihoods objectively favor the simple explanation *more* in the stochastic condition given this assumption—which, of course, works in the opposite direction of our hypotheses and our actual findings (where participants favored the simple explanation *less* in the stochastic condition). As a result, the current design almost certainly underestimates the real difference between conditions if the objective likelihoods are equated (as they were in Studies 2 and 5). Study S3 in the Supplementary Materials addresses this issue by directly comparing judgments about stochastic systems with specified (and equated) likelihoods versus vague (and unequated) likelihoods.

Study 4

A second contextual factor that may influence preferences of simple and complex explanations is the *content domain* of a causal system. People have abstract expectations about various domains such as the physical, biological, and social worlds (Wellman & Gelman, 1992). These can take the form of *overhypotheses* (Shipley, 1993) or *hierarchies of priors* (Kemp, Goodman, & Tenenbaum, 2010)—beliefs about what types of explanations are most plausible for such systems, irrespective of the concrete problem at hand.

Most importantly here, people believe that physical events have fewer causes than social events (Strickland et al., 2017) and appear to use causal concepts relying on physical transference for physical systems but complex counterfactual conditions for social systems (Lombrozo, 2010). Thus, people may favor simpler explanations in physical causal systems compared to social causal systems.

Electronic copy available at: https://ssrn.com/abstract=3194761

Studies 4 and 5 look at intuitions about simple and complex explanations across these domains, as well as biological systems and artifact systems. Specifically, Study 4 asks participants to judge the prior odds (Study 4A) and likelihood ratio (Study 4B) of simple versus complex explanations across these domains.

Methods. Participants (*N* = 240, 66 excluded) read 12 items across four content domains (physics, biology, artifact, and social). For example, one physics item read:

There is an array of ultraviolet waves radiating from the Arctic. Sometimes the waves display abnormal patterns such as frequency oscillations or irregular feedback.

Planck's effect can cause frequency oscillations.
A Bjork disturbance can cause irregular feedback.
The UV scatter effect can cause frequency oscillations.
The UV scatter effect can cause irregular feedback.

Nothing else is known to cause an ultraviolet wave to display frequency oscillations or irregular feedback.

Other physics items concerned subatomic particles and fluid dynamics. On the opposite end of the reductionist hierarchy, one of the social items read:

There is a volleyball tournament with many teams at a national gymnasium. Sometimes the teams have special strengths like good teamwork or positive reinforcement.

Mutual Trust can cause good teamwork.
Precision Leadership can cause positive reinforcement.
Collective Flourishing can cause good teamwork.
Collective Flourishing can cause positive reinforcement.

Nothing else is known to cause good teamwork or positive reinforcement in a volleyball team.

Other social items concerned child behavior and romantic attraction. The biological items concerned disease, agriculture, and dieting, while the artifact systems concerned robots, clocks, and toys.

Participants in Study 4A judged the prior odds of the simple and complex explanations for each item (using the same scale as Study 1A) and participants in Study 4B judged the likelihood ratio for each item (using the same scale as Study 1B). Items were completed in a random order.

Quantity	Physical	Biological	Artifact	Social
Prior Odds (Study 4A)	–1.11 (2.12)	–0.90 (1.96)	–0.75 (2.10)	0.08 (2.20)
Likelihood Ratio (Study 4B)	0.10 (2.25)	0.29 (2.45)	0.75 (2.32)	1.00 (2.19)

Table 4. Results of Study 4.

Note. Entries are relative probability judgments. Negative scores correspond to simple explanations, and positive scores to complex explanations. Scale ranges from –5 to 5. (SDs in parentheses.)

Electronic copy available at: https://ssrn.com/abstract=3194761

Results and discussion. As shown in Table 4, participants strongly favored the simple explanations for the physical items. Whereas Study 4A revealed a substantial bias favoring the simple explanation in assessing prior odds [$t(91) = 5.03$, $p < .001$, $d = 0.52$], Study 4B did not find a significant bias favoring the complex explanation in assessing likelihoods [$t(81) = 0.39$, $p = .70$, $d = 0.04$]. Thus, whereas participants in Study 1 had a complexity preference in assessing likelihoods, this was not the case for the physical causal systems used in Study 4.

The inferences for the social systems were the opposite as for the physical systems. Whereas Study 4B revealed a substantial bias favoring the *complex* explanation in assessing likelihoods [$t(81) = 4.13$, $p < .001$, $d = 0.48$], it did not find a significant bias favoring the simple explanation in assessing priors [$t(91) = 0.33$, $p = .74$, $d = 0.03$]. This led to a significantly stronger simplicity preference in the physical domain, compared to the social domain, both in assessing priors [$t(91) = 5.83$, $p < .001$, $d = 0.55$] and in assessing likelihoods [$t(81) = 3.90$, $p < .001$, $d = 0.41$].

These findings are consistent with the prediction, based on people's abstract causal theories of the physical and social world, that people favor simple, one-cause explanations for physical phenomena but are willing to entertain complex, multi-cause explanations for social phenomena. This is true for both pieces of the Bayesian posterior computation—people believe that simple explanations are more likely *a priori* in physical (but not social) systems, and people believe that complex explanations are better fits to the data in social (but not physical) systems.

The predictions for biological and artifact systems are less clear than for physical and social systems, but fall in between these reductionist extremes as one might expect (see Keil, Lockhart, & Schlegel, 2010 for a similar pattern in a different task). As shown in Table 4, the judgments for biological stimuli showed a substantial simplicity bias for prior odds [$t(91) = 4.38$, $p < .001$, $d = 0.46$] and a weaker, non-significant complexity bias for likelihoods [$t(81) = 1.08$, $p = .29$, $d = 0.12$]. This pattern is similar to the artificial biological (disease) items in Study 1, except the judgments were closer to the mean, likely due to the larger number and variety of items. Artifact stimuli split the difference between physical and social explanations, with moderately large biases both for prior odds [favoring simple explanations, $t(91) = 3.44$, $p < .001$, $d = 0.36$] and for likelihood ratios [favoring complex explanations, $t(81) = 2.91$, $p = .005$, $d = 0.32$]. This mixed result is consistent with the nature of artifacts—physical devices embedded in social contexts.

Although these results paint a consistent picture of simplicity and complexity preferences across domains, they are limited in asking only about the *components* of a Bayesian posterior calculation, rather than directly for explanatory inferences. Study 5 fills in this gap.

Study 5

Study 5 looks at both contextual moderators of explanatory judgments—determinism and domain—in the same design, measuring explanatory judgments. This allows us to test each moderator under broader conditions, as well as to look for possible boundary conditions.

Methods. Participants ($N = 479$, 89 excluded) read modified versions of the 12 items used in Study 4, with half of the participants judging the deterministic versions, such as:

Electronic copy available at: https://ssrn.com/abstract=3194761

There is an array of 750 ultraviolet waves radiating from the Arctic. You are a geoscientist trying to understand the abnormalities in the waves. One wave, the 185th, has two patterns: the 185th shows frequency oscillations and the 185th shows irregular feedback.

The **UV scatter** effect always (100% of the time) causes both frequency oscillations and irregular feedback.

Planck's effect always (100% of the time) causes frequency oscillations, but it never (0% of the time) causes irregular feedback.

A **Bjork disturbance** always (100% of the time) causes irregular feedback, but it never (0% of the time) causes frequency oscillations.

Nothing else is known to cause an ultraviolet wave to display frequency oscillations or irregular feedback.

The other half of participants judged the stochastic versions, which differed only in replacing "always" with "sometimes" and "100%" with "80%" (for the simple explanation) and "90%" (for the components of the complex explanation). Thus, these conditions are analogous to the 100% and 90% conditions from Study 2.

After reading this information, participants made explanatory judgments (e.g., "Which do you think is the most likely explanation for Wave 185's patterns?") on the same scale as Study 2. Items were completed in a random order.

Results and discussion. As shown in Table 5, the effects of both determinism and domain were as expected, given the theoretical framework and the results of the previous studies. First, participants favored the simple explanations more strongly for deterministic than for stochastic systems [$t(388) = 2.52$, $p = .012$, $d = 0.26$]. Thus, the shift seen in Study 2 was not unique to unfamiliar stimuli, or specific to reasoning about diseases. Rather, it is a much more general pattern that appears across many content domains.

Causal System	Physical	Biological	Artifact	Social
Deterministic	–2.76 (2.10)	–2.59 (2.19)	–2.32 (2.41)	–1.81 (2.71)
Stochastic	–2.15 (2.40)	–2.15 (2.28)	–1.81 (2.53)	–1.22 (2.59)

Table 5. Results of Study 5.

Note. Entries are explanatory preferences. Negative scores correspond to simple explanations, and positive scores to complex explanations. Scale ranges from –5 to 5. (SDs in parentheses.)

Second, the ordering of the means across domains was the same as in Study 4. Critically, participants had a stronger simplicity preference in the physical than in the social domain [$t(389) = 8.62$, $p < .001$, $d = 0.38$]. That is, not only do people make the *component* inferences (priors and likelihoods) in a manner favoring simple explanations more for physical than for social systems, but they also combine these inferences into posterior probabilities that favor simple explanations differentially across domains. Also like Study 4, the biological and artifact domains fell in between,

Electronic copy available at: https://ssrn.com/abstract=3194761

with the strongest preference for the physical, followed by the biological, artifact, then social domains.

These results complement those of previous studies, finding additive effects of both moderators on simplicity preferences. This helps to resolve the puzzle of how people could rely on a single cue—an explanation's simplicity—to do two logically independent jobs: estimating the prior and likelihood of an explanation, as well as combining these into an inference. If contextual moderators can influence the weighting of the simplicity and complexity heuristics, then a reasoner could reach different conclusions about simplicity and complexity in different contexts, in ways which are broadly adaptive.

However, despite the highly consistent ordinal effects across studies, there are lingering puzzles about what determines the strength and even direction of simplicity and complexity preferences. Even within the highly controlled studies presented here, these inferences appear to be subject to additional moderators.

On the one hand, one might have expected inferences to more strongly favor the simple explanations than they did here, given the strong simplicity preferences found for the artificial items in Study 2. The more moderate inferences here may have occurred because the items were seen as more reflective of the real world—where true determinism is rare—leading participants to hedge their bets. Alternatively, participants here could be recruiting background knowledge, relying more on memory rather than reasoning. In that case, the strong simplicity preferences found for artificial items in Studies 1 and 2 may actually be a better reflection of the underlying reasoning processes.

On the other hand, one might have expected some of the items—particularly in the social domain—to reveal an overall *complexity* preference. That is, since participants in Study 4 had a significant complexity preference in estimating likelihoods for social items, but no significant simplicity preference in estimating priors, those judgments should combine for a posterior favoring the complex explanations. One possibility is that the explicit probabilistic information given for the likelihoods—necessary to equate the normative explanatory judgments across the deterministic and stochastic conditions—has the side effect of making participants less likely to rely on a complexity heuristic for estimating likelihoods. The difference between the deterministic and stochastic systems—here as well as in Study 2—showed that the reliance on the heuristic can be modulated by the *nature* of probabilistic information, but we do not know to what extent the mere *presence* of probabilistic information attenuates its use. (See Study S3 for a more detailed examination of this issue.) In the real world, of course, events seldom wear probabilities on their sleeves, so natural conditions may favor the more even-handed use of the heuristics, compared to what we see in the current study.

One final issue is how confident we can be that differences across items in Studies 4 and 5 reflect differences across generalized expectations about the number of causes invoked across domains, versus idiosyncratic issues with our stimuli (e.g., differing expectations about the independence of causes across domains). Study S4 in the Supplementary Materials examines this question by measuring participants' general expectations for each of the items used in Studies 4 and 5 (e.g., "For a particular subatomic particle, do you think the particle's behavior is more likely to have a single cause or multiple causes?"). These expectations systematically differed across domains,

Electronic copy available at: https://ssrn.com/abstract=3194761

similar to Studies 4 and 5. In addition, mediation analyses revealed that these expectations influenced explanatory judgments because of their impacts on priors rather than on likelihoods, consistent with our account that these domain differences in explanation occur due to different overhypotheses across domains.

General Discussion

This paper set out to understand how people use simplicity to constrain their evaluation of theories or explanations, making tractable an otherwise ill-defined computational problem. Usually, simplicity is a good cue for an explanation's prior probability (intuitively, simple causes require fewer stars to align in order to occur) while complexity is a good cue for an explanation's likelihood or fit to the evidence (since complex causes have more opportunities to cause each aspect of the evidence). Study 1 found direct evidence for both of these *opponent heuristics*, directly asking about participants' priors and likelihoods.

However, our explanatory strategies must be definite enough to provide both a unique answer for a given explanatory problem, but also flexible enough to provide different answers to different problems. The opponent heuristics strategy solves this dilemma by modulating the inference depending on context. Study 2 showed that people shift toward complex explanations in stochastic contexts (because such contexts render a complexity heuristic more computationally relevant), and Study 3 suggested that these inferences in stochastic (and more ecologically realistic contexts) are probably more accurate. Studies 4 and 5 showed that people favor simple explanations to varying degrees across domains, in ways that track people's general expectations about the causal textures of these domains: People believe that physical systems are more linear, whereas social systems are more subject to branching, and people correspondingly favor simple explanations to a greater degree for physical systems.

Explanatory Logic. We view these opponent heuristics as one part of a broader *explanatory logic*—a set of heuristics and strategies that people use for evaluating explanatory hypotheses across a variety of psychological processes in light of our cognitive and informational limitations (Johnson, 2017; see also Lipton, 2001, 2004; McGrew, 2003). The current studies focused on causal explanation and our previous work has found similar effects in visual curve-fitting (Johnson, Jin, & Keil, 2014)—both tasks that require participants to evaluating competing hypotheses (causes, trend lines) for available data (effects, data points). However, many other cognitive processes also take this form, including categorization (which category best explains the features?), theory-of-mind (which mental state best explains the behaviors?), language (which meaning best explains the utterance?), and memory (which past events best explain the details I can recall?).

In ongoing work, we have been looking at simplicity heuristics in some of these other processes. For example, human beings often belong to a variety of social categories simultaneously—you can be a feminist bank teller, a Jewish woman, or a gay cognitive scientist. When explaining particular traits, people tend to favor social categorizations that invoke fewer categories simultaneously, but this bias is weaker when the categories are more loosely (i.e., stochastically) associated with the relevant features (Johnson, Kim, & Keil, 2016). Similarly, people favor mental-state explanations

Electronic copy available at: https://ssrn.com/abstract=3194761

that invoke relatively fewer goals to explain a particular behavior, but this simplicity preference is weaker when the goals are more stochastically associated with the behaviors (Johnson, 2017; Johnson, Hill, & Keil, 2016). Thus, opponent simplicity heuristics appear to pervade cognition.

Simplicity is not the only heuristic that appears to be widespread across different explanatory processes. People rely on inferences about unseen evidence (Johnson, Rajeev-Kumar, & Keil, 2016), find disconfirmed predictions more diagnostic than confirmed predictions (Johnson, Merchant, & Keil, 2015a), treat explanatory inferences as certainly true or false (Johnson, Merchant, & Keil, 2015b), and rely on diversity-based reasoning (Osherson et al., 1990) across a variety of psychological processes. We argue that these heuristics are, by and large, adaptive. For example, like simplicity, the tendency to make biased inferences about unseen evidence is used across widespread processes, including causal explanation (Khemlani, Sussman, & Oppenheimer, 2011), category-based induction (Johnson, Merchant, & Keil, 2015a), causal strength inferences (Johnson, Johnston, Toig, & Keil, 2014), social categorization (Johnson, Kim, & Keil, 2016), mental-state inference (Johnson, Hill, & Keil, 2016), and decision-making (Johnson, Zhang, & Keil, 2016). Although this heuristic can lead to errors, it is the product of a motivation to think about diagnostic evidence—a crucial way that humans are able to look beyond the observed. What, then, is the value of opponent simplicity heuristics?

The Adaptive Value of Opponent Heuristics. Our empirical argument for opponent heuristics has required us to engineer situations where people make errors. Nonetheless, we maintain that under more ecologically realistic conditions, these heuristics often serve us well and help to make explanatory reasoning possible.

If you have a well-specified prior distribution over the hypothesis space and you can construct a well-defined likelihood function, one arguably can do no better than normative Bayesian inference. Our participants fell well short of this standard, often making inferences unreasonably biased toward the simple explanations, which were influenced by factors that ought not have an influence, normatively speaking.

Some, however, have challenged Bayesian inference as a rational standard for evaluating hypotheses, instead advocating *explanationism*—the idea that explanatory considerations might impact posterior probability judgments over-and-above Bayesian calculations. For instance, one promising explanationist model simply adds an explanatory "bonus" to the Bayesian posterior for each hypothesis, proportional to how well that hypothesis explains the data (Douven, 2013). Descriptively, models of probability judgment that account for explanatory goodness outperform models that account only for Bayesian probabilities (Douven & Schupbach, 2015a, 2015b). Normatively, they may also outperform Bayesian updating in some respects. Explanationist models more rapidly converge to 1 for true hypotheses as evidence is accumulated (Douven, 2013), which may be useful for facilitating action in uncertain environments (Douven & Schupbach, 2015b; see also Johnson, Rajeev-Kumar, & Keil, 2015). Such models are known to perform especially well in social settings, where one's beliefs depend not only on direct evidence from the world, but also on others' beliefs (Douven & Wenmackers, 2017).

Electronic copy available at: https://ssrn.com/abstract=3194761

A further problem we face, even if we aspire to be Bayesian agents, is a lack of substantial information about probability distributions. We often are confronted with novel situations in which we cannot calculate but must simply guess, based on what little we can glean from the immediate problem and what minimal cues we can bring to bear from our previous experience. It may be true that people seldom encounter cases where they must diagnose an elf, deciding among unfamiliar diseases on the basis of make-believe symptoms, but it *is* true in real-world medical decision-making that we are often faced with highly limited information. Doctors have built up a corpus of statistical knowledge about some familiar diseases, and medical scientists may have some evidence to bring to bear on less familiar ones. Yet, *no one* has joint probability information about all combinations of diseases and symptoms. We must rely on iffy assumptions and fallible heuristics in order to make any real progress, even in a highly constrained problem domain such as medical diagnosis.

In other cases, probabilities may be even less evident. When making geopolitical forecasts, assessing the reasons for a friend's odd decision, or debating philosophical conundrums, there may be little relevant prior information, and it may be impossible to model the probabilities with any degree of confidence. This is known as *radical uncertainty* or *Knightian uncertainty* (Knight, 1921), and some philosophers hold that many sources of uncertainty are not quantifiable using probabilities (e.g., Levi, 1974; von Mises, 2008/1949). In cases of Knightian uncertainty, the best we can do is to adopt rules that work reasonably well most of the time. The use of simplicity and other explanatory heuristics appears to be one such adaptive habit.

This is not to claim that all bets are off, that our explanatory habits are untethered to the world. On the contrary, simplicity is usually an excellent principle to use for assessing explanations, because there are often multiple explanations, varying in complexity, which fit the data equally well. In such cases, the priors often *do* favor simple explanations, so a simplicity heuristic is reasonable. But when the explanations vary in likelihood, simplicity will lead us astray, as complex explanations are often better fits to the data. An opponent heuristic system allows us to harness both of these general facts about the world to our cognitive advantage, while avoiding complex computations that may be intractable and, in Knightian cases, even impossible.

Electronic copy available at: https://ssrn.com/abstract=3194761

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*, 1156–1164.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22.

Douven, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimization. *Philosophical Quarterly*, *63*, 428–444.

Douven, I., & Schupbach, J. N. (2015a). The role of explanatory considerations in updating. *Cognition*, *142*, 299– 311.

Douven, I., & Schupbach, J. N. (2015b). Probabilistic alternatives to Bayesianism: The case of explanationism. *Frontiers in Psychology, 6*(459).

Douven, I., & Wenmackers, S. (2017). Inference to the best explanation versus Bayes's rule in a social setting. *The British Journal for the Philosophy of Science*, *68*, 535–570.

Dowe, P. (2000). *Physical causation*. Cambridge, UK: Cambridge University Press.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–633.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less *ad hoc* theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, *45*, 1–35.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1–32.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford.

Hochberg, J., & McAlister, E. (1953). A quantitative approach to figural "goodness." *Journal of Experimental Psychology*, *46*, 361–364.

Hood, B. M. (1995). Gravity rules for 2- to 4-year olds? *Cognitive Development*, *10*, 577–598.

Johnson, S. G. B. (2017). *Cognition as sense-making*. Unpublished doctoral dissertation.

Johnson, S. G. B., Hill, F., & Keil, F. C. (2016). *Explanatory heuristics in theory-of-mind*. Working paper.

Johnson, S. G. B., Jin, A., & Keil, F. C. (2014). Simplicity and goodness-of-fit in explanation: The case of intuitive curve-fitting. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 701–706). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Johnston, A. M., Toig, A. E., & Keil, F. C. (2014). Explanatory scope informs causal strength inferences. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2453–2458). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Kim, H. S., & Keil, F. C. (2016). Explanatory biases in social categorization. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 776–781). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Merchant, T., & Keil, F. C. (2015a). Argument scope in inductive reasoning: Evidence for an abductive account of induction. In D. C. Noelle, R. Dale, A. S. Warlaumont, J.

Electronic copy available at: https://ssrn.com/abstract=3194761

Yoshimi, T. Matlock, C. D. Jennings, & P.P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1015–1020). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Merchant, T., & Keil, F. C. (2015b). Predictions from uncertain beliefs. In D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, & P.P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1003–1008). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2015). Belief utility as an explanatory virtue. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1009–1014). Austin, TX: Cognitive Science Society.

Johnson, S. G. B., Rajeev-Kumar, G., & Keil, F. C. (2016). Sense-making under ignorance. *Cognitive Psychology*, *89*, 39–70.

Johnson, S. G. B., & Rips, L. J. (2015). Do the right thing: The assumption of optimality in lay decision theory and causal judgment. *Cognitive Psychology, 77*, 42–76.

Johnson, S. G. B., Zhang, M., & Keil, F. C. (2016). Decision-making and biases in causal-explanatory reasoning. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1967–1972). Austin, TX: Cognitive Science Society.

Keil, F. C., Lockhart, K. L., & Schlegel, E. (2010). A bump on a bump? Emerging intuitions concerning the relative difficulty of the sciences. *Journal of Experimental Psychology: General*, *139*, 1–15.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*, 1185–1243.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2011). *Harry Potter* and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*, 527–535.

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Boston, MA: Hart, Schaffner, & Marx.

Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhya, Series A, 25*, 369–376.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, *71*, 391–418.

Lewis, D. (2000). Causation as influence. *The Journal of Philosophy*, *97*, 182–197.

Lipton, P. (2000). Is explanation a guide to inference? A reply to Wesley C. Salmon. In G. Hon and S. S. Rakover (Eds.), *Explanation: Theoretical approaches and applications* (pp. 93–120). Dordrecht, Netherlands: Springer.

Lipton, P. (2004). *Inference to the best explanation* (2nd Ed.). London: Routledge.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*, 232–257.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*, 303–332.

Mackie, J. L. (1965). Causes and conditions. *American Philosphical Quarterly*, *2*, 245–264.

McGrew, T. (2003). Confirmation, heuristics, and explanatory reasoning. *The British Journal for the Philosophy of Science*, *54*, 553–567.

Electronic copy available at: https://ssrn.com/abstract=3194761

Osherson, D., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.

Pacer, M., & Lombrozo, T. (2017). Occam's razor cuts to the root: Simplicity in causal explanation. *Journal of Experimental Psychology: General*, *146*, 1761–1780.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, *26*, 303–343.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*, 465–471.

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), *The psychology of learning and motivation: Vol. 30* (pp. 265–301). San Diego, CA: Academic Press.

Sober, E. (1983). *Reconstructing the past: Parsimony, evolution, and inference*. Cambridge, MA: MIT Press.

Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development*, *60*, 637–648.

Strickland, B., Silver, I., & Keil, F. C. (2017). The texture of causal construals: Domain-specific biases shape causal inferences from discourse. *Memory & Cognition*, *3*, 442–455.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*, 435–502.

Van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review*, *163*, 429–456.

Von Mises, L. (2008). *Human action: A treatise on economics*. Auburn, AL: Ludwig von Mises Institute. (Original work published 1949.)

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337–375.

Wilde, O. (2004). *The importance of being earnest*. New York, NY: Penguin. (Original work first performed 1895.)

Electronic copy available at: https://ssrn.com/abstract=3194761

Supplementary Materials

Study S1

A possible concern about the method used in Study 1 to elicit simplicity and complexity preferences for priors and likelihoods is that such preferences may depend on *what* is being explained, as explanatory goodness is a link between a putatively explanatory hypothesis and a particular observation (e.g., Lipton, 2000). In Study 1, judgments were not made in reference to a particular event that needed explaining (e.g., Wenlie's symptoms) but rather the probabilistic features of diseases in the population at large. People may well reason differently when making judgments about individuals versus populations, and when making judgments in explanatory versus non-explanatory contexts. Thus, Study S1 sought to replicate the basic finding of Study 1 while asking participants to make their judgments in reference to an individual rather than a group. In addition, Study S1 asked participants to rate the priors and likelihoods separately for the simple and complex explanations, as opposed to Study 1's method of asking for relative judgments on a single scale.

Methods. Participants (N = 240, 67 excluded) were randomly assigned to Study S1A (judgments about prior probabilities) or to Study S1B (judgments about likelihoods). The items were similar to those used in Study 1 from the main text, except that a specific individual was mentioned in explaining the diseases and symptoms (e.g., "There is a population of elves that lives at Gelfert's Glacier. One elf, Wenlie, has feverish muffets and wrinkled ears.").

Participants in Study S1A were then asked to judge the prior probabilities, first being instructed to "Suppose that you did not know about Wenlie's symptoms—that is, suppose that Wenlie were just a randomly selected elf from Gelfert's Glacier." They then were asked to separately rate the priors of the individual having each having the simple and complex set of diseases on 0–10 scales. Finally, participants decided, as a forced choice, whether it was likely that there were more individuals who had the simple or complex set of diseases.

The procedure was similar for Study S1B, except they were asked to "Suppose you did not know about Wenlie's symptoms, but you knew for sure that Wenlie had [a Yewlie infection only / both Hepz's disease and Aeona's syndrome]," with the simple or complex set of diseases filled in for rating the corresponding likelihood. Then participants were asked to rate the likelihoods ("How likely is it that Wenlie would develop both feverish muffets and wrinkled ears") separately for the simple and complex diseases on the same scale as Study S1A. Participants also decided, as a forced choice, whether elves with the simple or the complex set of diseases were likelier to develop both symptoms.

As in the other studies, participants were excluded if they incorrectly answered more than 30% of a series of check questions. Participants were also eliminated if their forced-choice response was inconsistent with their numerical ratings (i.e., rating the simpler explanation higher in the numerical ratings but choosing the complex explanation, or vice versa).

Electronic copy available at: https://ssrn.com/abstract=3194761

Results and discussion. The results were highly consistent with those from Study 1 (see Table S1). The prior odds were estimated to be much higher for the simple than for the complex explanation [$t(89) = 7.94$, $p < .001$, $d = 0.84$], whereas the likelihood of the data were estimated as somewhat higher for the complex than for the simple explanation [$t(82) = 2.08$, $p = .041$, $d = 0.23$].

Quantity	Simple	Complex
Prior Odds (Study S1A)	5.53 (2.53)	2.95 (2.22)
Likelihood Ratio (Study S1B)	7.10 (1.84)	7.51 (1.95)

Table S1. Results of Study S1.

Note. Entries are likelihood judgments on a scale from 0 to 10. (SDs in parentheses.)

These results are very similar to those of Study 1, with a large simplicity preference for judging priors and a more modest complexity preference when judging likelihoods. Thus, these intuitions are robust to differences in the measurement scale (separate versus relative judgments) and to specification of an individual versus a population (e.g., judging the prior probability and likelihood for Wenlie the elf as opposed to elves in general).

Study S2

The shift in simplicity preferences between deterministic and stochastic systems need not be non-normative, and several assumptions must be satisfied if these conditions are to be normatively equivalent—in particular, the prior probabilities need to be the same across the deterministic and stochastic conditions, and the likelihoods need to be the same across conditions (this requires the effects to be independent, conditional on their causes). In Study S2, we devised materials that ensured that these assumptions are met, while extensively checking that participants shared these assumptions with us.

Methods. Participants ($N = 484$, 69 excluded) were randomly assigned to Study S2A (deterministic causal systems) or to Study S2B (stochastic causal systems).

Participants read materials based on the elf cover story from Studies 1–3:

> Below is some information about diseases in elves that can cause feverish muffets and wrinkled ears.
>
> **Yewlie infection**
> A Yewlie infection can cause an elf to get feverish muffets and wrinkled ears. A Yewlie infection causes feverish muffets and wrinkled ears in [100% / 80%] of cases. The chance that a Yewlie infection causes an elf to develop these symptoms is the same regardless of whether the elf also has other diseases.
>
> **Hepz's disease**
> Hepz's disease can cause an elf to get feverish muffets (but never causes wrinkled ears). Hepz's disease causes feverish muffets in [100% / 90%] of cases. The chance that Hepz's disease causes an elf to develop feverish muffets is the same regardless of whether the elf also has other diseases.

Electronic copy available at: https://ssrn.com/abstract=3194761

Aeona's syndrome
Aeona's syndrome can cause an elf to get wrinkled ears (but never causes feverish muffets). Aeona's syndrome causes wrinkled ears in [100% / 90%] of cases. The chance that Aeona's syndrome causes an elf to develop wrinkled ears is the same regardless of whether the elf also has other diseases.

These three diseases occur equally often among elves.

For these three diseases, having one disease does not make an elf more or less likely to come down with another disease.

Nothing else is known to cause an elf's muffets to be feverish or the development of wrinkled ears.

Participants were then asked a series of multiple-choice questions on the same screen, probing the likelihoods of the symptoms given the simple and complex diseases, the independence of the causes, the conditional independence of the effects, the equality of the base rates of the diseases, and the (non-)existence of alternative causes. Participants incorrectly answering any of these questions were asked to carefully reread the materials and to answer the questions again. Any participant who did not correctly answer all six questions after two attempts was excluded from the primary data analysis ($N = 68$), although the results are nearly identical if these participants are included.

On the next screen, participants were told that "Wenlie is an elf who has both feverish muffets and wrinkled ears," and asked to diagnose what is causing Wenlie's symptoms. Like Study 3, participants rated the probability of each explanation separately ("Please estimate the probability of Wenlie having each combination of diseases" for "a Yewlie infection" and "both Hepz's disease and Aeona's syndrome") on separate scales from 0 to 100. Unlike Study 3, participants were not asked to make the probabilities sum to 100.

Results and discussion. The results were similar to the results of Studies 2 and 3. The simple explanation was more strongly favored given the deterministic than the stochastic causal system [$t(413) = 2.68$, $p = .008$, $d = 0.26$]. However, preferences for the complex explanation did not significantly differ across conditions [$t(413) = 0.39$, $p = .70$, $d = 0.04$].

	Posterior Probabilities	
Explanation	**Deterministic (Study S2A)**	**Stochastic (Study S2B)**
1-cause	76.5% (23.2%)	70.5% (22.2%)
2-cause	51.2% (30.9%)	50.1% (29.9%)

Table S2. Results of Study S2.

Note. Entries for the Deterministic and Stochastic columns are judged posterior probabilities, expressed as percentages (SDs in parentheses).

This latter null result is difficult to interpret in isolation, as the normative claims we can make are about the *ratio* of the posteriors of the simple and complex explanations. For instance, if participants

Electronic copy available at: https://ssrn.com/abstract=3194761

believed that deterministic causes had higher base rates than stochastic causes, then they might rationally assign higher posterior probabilities to both the 1-cause *and* the 2-cause explanation. Since we specified only that the base rates of the component causes are equal but not their magnitudes, all we can say is that it is non-normative to assign a higher posterior to the 1-cause but not to the 2-cause explanation, and that this non-normative behavior is consistent with the heuristics we have posited.

Study S3

When supplying likelihoods for stochastic causal systems, we face a choice between two imperfect options: Making the likelihoods precise (e.g., the cause leading to the effect 90% of the time) or leaving them vague (e.g., the cause "sometimes" leading to the effect). We used precise likelihoods in most of our studies so far that included this manipulation (Studies 2, 5, and S2) because this choice has two important advantages. First, it avoids confounding precision with stochasticity (e.g., "always" equals exactly 100% but "sometimes" does not equal any precise probability). Second, and most critically, it allows us to equate the normative likelihood of the combined complex explanation. If participants assume that the simple cause leads to both of its effects 80% of the time, while each component of the complex cause leads to its particular effect 80% of the time, then the complex cause leads to both effects together only 65% of the time (using the assumptions checked in Study S2). This makes the simple explanation *normatively* more probable in the stochastic case, even though this is not the case in the deterministic case. Thus, we have a confound that works against our hypothesis and tends to attenuate any complexity preference found in the deterministic case.

On the other hand, using precise likelihoods has a distinct disadvantage: There is less room for a complexity heuristic to operate. We predicted a weaker preference for complex explanations under deterministic conditions because it is easy to conceptualize the likelihoods in deterministic cases (being 100%), whereas this is more difficult in stochastic cases (being somewhere between 0% and 100%). Complexity can be used as a way to indirectly get a handle on the likelihood. But why should we even *need* an indirect handle on the likelihood when we have the likelihood's precise value (e.g., 90%)?

In fact, we think there are multiple reasons why humans have difficulty working with such probabilities directly. First, people are known to think about probabilities in a "digital" rather than "analog" way (Johnson, Merchant, & Keil, 2015b; Murphy & Ross, 1994), such that we would expect people to avoid reasoning about precise probabilities whenever possible. Second, as a more general point, probabilities of 0 and 1 often lead to computational savings because they often fall out in the normative computations. For instance, when calculating likelihood ratios, the ratio is trivial to compute when both likelihoods are 1, but is not trivial when either likelihood differs from 1. Particularly in comparing simple and complex hypotheses, the *only* situation in which multiplication or division is not required is when the likelihoods are deterministic. In our experimental set-up, the simple stochastic cause has a likelihood (80%) that happens to be about equal to the product of the two components of the complex cause (90% x 90% ≅ 80%). While it is easy to divide 80% by 80%, it is effortful to compute the denominator of this ratio (which only happens to be 80%).

Electronic copy available at: https://ssrn.com/abstract=3194761

The take-away from all this is that there are two competing forces in considering whether vague or precise probabilities would lead to a stronger simplicity preference: Heuristic forces that may render a complexity heuristic more useful for vague probabilities, leading to a stronger complexity preference; and normative forces that render the complex explanation less objectively likely for vague probabilities (given our experimental set-up). Such a situation calls for an experimental test. Study S3 compared deterministic systems, stochastic systems with precise likelihoods, and stochastic systems with vague likelihoods. We did so in two separate experiments, one using separate scales measuring the probability of each explanation and one using a single scale measuring the relative likelihood of the explanations.

Methods. Participants (*N* = 318, 35 excluded) participated in Study S3A (separate probability scales) or Study S3B (bipolar relative likelihood scale).

Participants each completed three items, one with deterministic likelihoods, one with precise stochastic likelihoods, and one with vague stochastic likelihoods. These were based on three of the cover stories used in Studies 1–3. The materials for the elf story read:

> Below is some information about diseases in elves that can cause feverish muffets and wrinkled ears.
>
> **Yewlie infection**
> A Yewlie infection can cause an elf to get feverish muffets and wrinkled ears. A Yewlie infection causes feverish muffets and wrinkled ears in [*100% of / 80% of / many, but not all,*] cases. The chance that a Yewlie infection causes an elf to develop these symptoms is the same regardless of whether the elf also has other diseases.
>
> **Hepz's disease**
> Hepz's disease can cause an elf to get feverish muffets (but never causes wrinkled ears). Hepz's disease causes feverish muffets in [*100% of / 90% of / many, but not all,*] cases. The chance that Hepz's disease causes an elf to develop feverish muffets is the same regardless of whether the elf also has other diseases.
>
> **Aeona's syndrome**
> Aeona's syndrome can cause an elf to get wrinkled ears (but never causes feverish muffets). Aeona's syndrome causes wrinkled ears in [*100% of / 90% of / many, but not all,*] cases. The chance that Aeona's syndrome causes an elf to develop wrinkled ears is the same regardless of whether the elf also has other diseases.
>
> These three diseases are rare and occur equally often among elves.
>
> For these three diseases, having one disease does not make an elf more or less likely to come down with another disease.
>
> Nothing else is known to cause an elf's muffets to be feverish or the development of wrinkled ears.

On the bottom of the screen, participants were told that "Wenlie is an elf who has both feverish muffets and wrinkled ears," and asked to diagnose what is causing Wenlie's symptoms. In Study S3A (like Studies 3 and S2A), participants rated the probability of each explanation separately ("Please estimate the probability of Wenlie having each combination of diseases" for "a Yewlie infection

Electronic copy available at: https://ssrn.com/abstract=3194761

only" and "Hepz's disease and Aeona's syndrome only") on separate scales from 0 to 100. Participants were not asked to make the probabilities sum to 100. In Study S3B (like Studies 2 and 4), participants rated the relative likelihood of the diseases ("Which do you think is the most likely explanation for Wenlie's symptoms?") on a scale from –5 ("Wenlie has a Yewlie infection") to 5 ("Wenlie has Hepz's disease and Aeona's syndrome").

Results and discussion. As shown in Table S3, the core result from previous studies—the stronger preference for simple explanations in deterministic causal systems—was consistently found across both Studies S3A and S3B. The difference between precise and vague stochastic systems, however, was less consistent across studies and appears to depend on the particulars of the task.

When posterior probabilities were measured on separate scales in Study S3A, participants rated the 1-cause explanation as more probable in the deterministic than in the stochastic/precise condition [$t(137) = 4.58$, $p < .001$, $d = 0.29$] and the 2-cause explanation as more probable in the stochastic/precise than in the deterministic condition [$t(137) = 3.18$, $p = .002$, $d = 0.23$]. This replicates the results of Studies 2, 5, and S2A.

Posterior Probabilities			
Explanation	**Deterministic**	**Stochastic (Precise)**	**Stochastic (Vague)**
1-cause (Study S3A)	79.9% (21.9%)	73.5% (21.4%)	71.5% (20.0%)
2-cause (Study S3A)	38.1% (28.9%)	44.8% (30.9%)	35.2% (23.7%)
Relative Preference (Study S3B)	–3.57 (2.03)	–2.00 (3.31)	–2.90 (2.47)

Table S3. Results of Study S3.

Note. Entries for the first two rows are judged posterior probabilities, expressed as percentages. Entries for the third row are explanatory judgments. Negative scores correspond to simple explanations, and positive scores to complex explanations. Scale ranges from –5 to 5. (SDs in parentheses).

The results were less clear for the comparison between the stochastic/precise condition and stochastic/vague conditions. The posterior assigned to the 1-cause explanation was numerically lower than in the stochastic/vague compared to the stochastic/precise condition, in keeping with the idea that vagueness adds an additional reason to rely on the complexity heuristic, but this difference did not reach significance [$t(137) = 1.30$, $p = .19$, $d = 0.09$]. The posterior assigned to the 2-cause explanation was actually lower in the stochastic/vague condition [$t(137) = 4.53$, $p < .001$, $d = 0.35$], in keeping with the idea that the complex explanation *is* normatively less likely in the vague condition. Thus, Study S3A does not provide convincing evidence that people differentiate in a consistent way between vague and precise versions of stochastic systems.

The results for Study S3B were cleaner. Once again, participants sharply differentiated between the deterministic and stochastic/precise condition, with a dramatically stronger simplicity preference in the deterministic condition [$t(144) = 6.40$, $p < .001$, $d = 0.59$], consistent with the earlier results.

Electronic copy available at: https://ssrn.com/abstract=3194761

This time, however, the results clearly fell in the middle for the stochastic/vague condition, with the simplicity preference weaker than in the deterministic condition [$t(144) = 3.00$, $p = .003$, $d = 0.30$] and stronger than in the stochastic/precise condition [$t(144) = 3.54$, $p < .001$, $d = 0.31$]. This supports the idea that the normative forces overwhelm the heuristic forces for the particular trade-off we are looking at here. Of course, the large difference between the deterministic and stochastic/precise condition, in both Studies S3A and S3B, shows that heuristic forces matter greatly, as there is no normative difference between these two conditions.

On balance, these results strongly and consistently replicate differences between deterministic and (precisely defined) stochastic systems. There are two competing forces that would tend to lead to different results vis-à-vis precisely versus vaguely stochastic systems. Although the results were not entirely consistent across studies, the greater weight of evidence points to the normatively higher posteriors for simple explanations in the vague case overwhelming the heuristic factors that might lead to a weaker simplicity preference in the vague case. These normative factors are not, however, powerful enough to overwhelm the generally much stronger simplicity preference for deterministic systems.

Given the sensitivity of this result to the task (separate scales versus one bipolar scale), it seems likely that other experimental set-ups could reverse the stronger complexity preference for vague likelihoods. For example, factors that decreased the task demand to assign similar likelihoods to all three component explanations (e.g., using *different* vague probability words for all three likelihoods rather than consistently using "sometimes" or "often") could decrease the pressure of the normative force, unleashing the heuristic force in its full potency.

Study S4

Our hypotheses about simplicity and complexity preferences across content domain—tested in Studies 4 and 5—were motivated by the idea that people have general expectations (overhypotheses; Kemp et al. 2010; Shipley, 1993) about the causal structure of different domains. For example, prior research has found that people believe physical events to have fewer causes than social events in general (Strickland et al., 2017). We predicted that these general expectations about content domains would translate into inferences about specific problems.

Arguably, however, there is a link missing in this chain of evidence. Although we know that people have general expectations about causation across domains (Strickland et al., 2017) and we know that people favored different explanations across different domains (Studies 4 and 5), we have not shown that general domain-driven expectations were applied to the specific stimuli we studied. Absent such evidence, it remains possible that domain differences across our stimuli were confounded with other differences (e.g., in the independence of the causes). To test this, we asked participants about their general causal beliefs, concerning the number of causes likely to cause the 12 types of events studied in Studies 4 and 5.

Methods. Participants ($N = 80$, 7 excluded) participated in Study S4. Participants completed 12 items, each corresponding to one of the items used in Studies 4 and 5 from the physical, biological, artifact, and social domains. For example, one physical item read "For a particular subatomic

Electronic copy available at: https://ssrn.com/abstract=3194761

particle, do you think the particle's behavior is more likely to have a single cause or multiple causes?" and one social item read "For a particular volleyball team, do you think their success is more likely to have a single cause or multiple causes?" These questions were answered on scales from –5 ("a single cause") to 5 ("multiple causes").

Results and discussion. As shown in Table S4, expectations about the number of causes across content domains largely tracked the explanatory preferences found in Studies 4 and 5. Mirroring previous studies, participants were much more likely to favor multiple causes in the social domain compared to the physical domain [$t(72) = 11.54$, $p < .001$, $d = 1.89$]. The domains follow the same order as in Studies 4 and 5, except that artifact-related events were seen as likely to have single causes in Study S4, whereas artifacts fell between the biological and social domains in Studies 4 and 5. One possible explanation for this discrepancy is that participants are likelier to appreciate the mechanistic complexity of artifacts when confronted with more specified information about the artifact's components (Rozenblit & Keil, 2002), as they were in Studies 4 and 5 but not Study S4.

Physical	Biological	Artifact	Social
–0.26 (1.83)	1.62 (1.86)	–0.79 (2.10)	2.97 (1.57)

Table S4. Results of Study S4.

Note. Entries are beliefs about the likely number of causes for each problem. Negative scores correspond to a single cause, and positive scores to multiple causes. Scale ranges from –5 to 5. (SDs in parentheses.)

Our conceptual model would predict that differences in general causal expectations would lead to differences in prior probabilities (i.e., overhypotheses), which in turn would lead to differences in explanatory preferences. To test this idea, we conducted a mediation analysis, using the item means for each of the 12 items (averaged across participants), for our measures of general causal expectations (Study S4), priors (Study 4A), likelihoods (Study 4B), and explanatory preferences (Study 5, averaging across the deterministic and stochastic conditions). We fit a multiple mediation model (Model 6 in the PROCESS macro; Hayes, 2013), with general causal expectations as the predictor, likelihoods and priors as potential mediators (in that order), and explanatory preferences as the outcome.

General causal expectations were not significant predictors of likelihoods, 95% CI[–0.08,0.25], but were significant marginally predictors of priors (adjusting for likelihoods), 95% CI[–.01,0.27]. When priors, likelihoods, and general causal expectations were entered simultaneously into a regression to predict explanatory preferences, both priors, 95% CI[0.13,0.83], and likelihoods, 95% CI[0.03,0.85], had significant impacts on explanatory preferences, but general causal expectations had no further impact beyond these potential mediators, 95% CI[–0.09,0.07], indicating that any mediating effect is full. Finally, tests for indirect effects revealed a significant indirect effect via priors (Expectations -> Priors -> Explanations), 95% CI[0.00,0.14], but not via likelihoods (Expectations -> Likelihoods -> Explanations), 95% CI[–0.05,0.16], nor the serial mediation pathway (Expectations -> Likelihoods -> Priors -> Explanations), 95% CI[–0.03,0.19]. The results

Electronic copy available at: https://ssrn.com/abstract=3194761

are similar if the mediators are entered in the opposite order (a marginal result becomes significant and a significant result becomes marginal), and the results are similar if the mediators are entered individually in separate models (i.e., PROCESS Model 4).

Taken together, these mediation results show that general causal expectations, as measured in Study S4, explain the domain differences found in Studies 4 and 5. Moreover, the effect of domain on explanatory preferences occurred because different domains are associated with different assignments of prior probability to different causal hypotheses, rather than because of differences in inferred likelihoods. Overall, this pattern of results strongly supports our conceptual model.

Electronic copy available at: https://ssrn.com/abstract=3194761